DL

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 February 2001 (01.02.2001)

PCT

(10) International Publication Number
WO 01/07920 A2

(51) International Patent Classification[7]: G01N 33/68

(21) International Application Number: PCT/US00/20268

(22) International Filing Date: 25 July 2000 (25.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/361,789 27 July 1999 (27.07.1999) US

(71) Applicant: BIO-RAD LABORATORIES, INC. [US/US]; 1000 Alfred Nobel Drive, Hercules, CA 94547 (US).

(72) Inventors: NGUYEN, Quan; 288 Whitmore No. 124, Oakland, CA 94611 (US). RANDALL, Jon; 323 Elati Court, Danville, CA 94526 (US). DAVIS, Diana; 880 Mission Way, Sacramento, CA 95819 (US).

(74) Agents: HEINES, M., Henry et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th Floor, San Francisco, CA 94111-3834 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/07920 A2

(54) Title: AUTOMATED LANDMARKING FOR TWO-DIMENSIONAL ELECTROPHORESIS

(57) Abstract: Two-dimensional electrophoresis gels are analyzed in sets, each set containing a multitude of gels where the gels typically represent samples taken at different times from the same source. This type of analysis permits observations of sequential progressions and changes in the protein compositions of biological samples. The present disclosure provides a means of standardizing, normalizing, and establishing proper correspondence between the protein spots in the various gels in the set, by the addition of marker proteins whose spots are detectable in a manner that distinguishes them from the spots representing the proteins of the samples. The gels are therefore scanned twice, once to detect the marker protein spots and once to detect the sample protein spots, the former serving as landmarks for the automated processing of the scanning data. This avoids the need for manual landmarking as practiced in the corresponding systems of the prior art.

AUTOMATED LANDMARKING FOR TWO-DIMENSIONAL ELECTROPHORESIS

This invention relates to areas of technology involving two-dimensional electrophoresis, imaging of two-dimensional electropherograms, identification and quantitation of protein spots or bands in electrophoresis, and the compilation and processing of data in series of multiple two-dimensional electrophoresis gels.

BACKGROUND OF THE INVENTION

Two-dimensional electrophoresis is widely used as a means of identifying, monitoring and quantifying the compositions of complex biological mixtures. Separations performed in two dimensions enable the biochemist to detect and distinguish a large number of components that would not be distinguishable in a linear separation. While the identification and analysis of the separated components in a two-dimensional electrophoresis gel is a complex and difficult task, methods have been developed to simplify, standardize and automate the procedure.

Among these methods are those designed for analyzing and comparing each gel in a series of two-dimensional electropherosis gels. This finds application in situations where each gel in the series represents a successive stage of a clinical, diagnostic or analytical procedure or successive stages a life cycle. The variations and progressions from one stage to the next serve to provide information that is valuable for a number of reasons. Certain biological cells, for example, have life cycles that are defined by changes in their morphology and physiology, such as changes in the expression of particular genes. The developmental stages of these cells can be defined and monitored by their global patterns of gene expression and the progressive changes that occur in these global patterns and in particular genes. Accelerated cell growth, reductions in cell growth, mutations and other transformations, patterns and processes can be monitored as a means of detecting and monitoring the response of the cell to chemical factors such as drug and hormone treatments, nutrient availability, and other growth conditions, and environmental factors such as heat shock, pH, and other stresses. Tumor growth can be monitored by monitoring the DNA content of a cell. Information of this type can be used to classify tumors in terms of high or low malignancy, therapies can be evaluated and

adjusted, and other types of information useful in the diagnosis, treatment and prognosis of cancers can be obtained. As further examples, cells in which genes have been inserted can be monitored to address questions such as whether or not the gene is expressing a particular protein, whether or not the cell's progeny is expressing the protein, whether the protein is being secreted, and whether mutations in the gene have occurred. These are examples of the many types of information that can be obtained by analyses of multiple two-dimensional electrophoresis gels taken in sequence. Other examples are well known among biochemists.

The manner in which the spots are distributed across a two-dimensional gel will depend on the separation parameters used in the first and second dimension stages of the separation. Two parameters that are commonly used in these separations are isoelectric point and molecular weight, although any parameter that can serve as the basis for an electrophoretic separation can be used. Whatever the parameter(s), the number of spots in a two-dimensional gel may be as high as several hundred, and each is detected by means of a signal derived from the individual spot. Various types of signals have been used, examples of which are optical density, radioactive emissions, fluorescence emissions, and colorimetric signals. Detection of the spots is commonly achieved by the use of imaging devices that convert these signals into digital data that is electronically readable data. The conversion is performed by imaging systems that store the information on computer storage media. The data typically stored are data indicating molecular weight, isoelectric point and quantity or concentration. Once stored, the data is processed by the computer in a manner that presents the data in a progressive manner, for example by showing the appearance, disappearance, or growth profile of a particular protein spot throughout the course of the time spanned by the gels. A variety of software programs that perform these functions are commercially available to biochemical laboratories. Examples of such programs are "PDQUEST" and "Melanie" (both available from Bio-Rad Laboratories, Inc., Hercules, California, USA), and "ImageMaster 2-D" (available from Pharmacia Biotech, Cleveland, Ohio, USA),.

Groups of gels that are to be compared to each other in a qualitative and quantitative manner for purposes of observing progressions and other changes in individual protein spots are commonly referred to by the software manufacturers and users as "matchsets." The software currently in use calls for the inclusion within each matchset of a "standard" that is constructed by the user. The standard is constructed from several electropherograms within the matchset that are selected to collectively contain all

of the spots that appear in any of the gels of the matchset. With all of the spots included and identified, the standard serves to identify particular protein spots in any single gel and to locate corresponding spots for particular proteins in different gels.

One of the key features of the standard is the assignment and use of reference spots known as "landmarks." The landmarks are actual protein spots that are manually selected by the user, the software then processing the manually selected spots to automatically identify the same spots in all of the member gels of the matchset. These user-selected landmarks are relatively few in number compared to the total number of spots in an individual electropherogram. The selection criteria are that the spots be well-resolved, that they be isolated from other spots, and that they appear in all gels of the matchset. The number of spots to be assigned as landmarks must be large enough that all of the remaining protein spots among the various gels will be successfully matched by the automatic processing. The function of the landmarks is to serve as guideposts in the gel-to-gel comparisons, compensating and correcting for slight differences and distortions among the member gels in the matchset to assure that there will be a proper correspondence of protein spots among different gels in the matchset.

User guides that accompany the software packages set forth the criteria that the user should apply in selecting the spots to be used as landmarks and the procedures that the user must follow in assigning landmark status to those spots. Users find however that the process of selecting and marking spots to be used as landmarks is slow and tedious and is the limiting factor in two-dimensional data analysis. The need to adhere to the selection criteria enumerated above and the time involved in making the selections and marking them adds to the cost and time involved in performing the analyses, and the level of user involvement raises questions of reproducibility and reliability.

SUMMARY OF THE INVENTION

These and other disadvantages and limitations of manual landmarking in the analyses of series of two-dimensional electropherograms are addressed by the present invention, in which manual landmarking is replaced by the introduction of marker proteins into the electrophoretic separations. Two-dimensional separations are performed on the marker proteins and the sample proteins together to form two-dimensional arrays of protein spots, each such array including both marker protein spots and sample protein

spots. The marker proteins are labeled in a manner that is distinguishable, *i.e.*, separately detectable, from the sample proteins, and are identified in the scanned images as marker proteins by these distinguishing labels rather than by markings applied to the scanned images as is done in the prior art. Each marker protein is characterized by a known value of the same parameter or parameters that are used for identifying the sample proteins. Thus, in the frequently used technique of two-dimensional electrophoresis where the location of each sample protein in the array is determined by the molecular weight and isoelectric point of that protein, the molecular weight and isoelectric point of each marker protein will be known prior to their use in the analysis. Other parameters on which electrophoretic separations can be based can be substituted for molecular weight, isoelectric point or both. In any case, once the separations have been performed and the two-dimensional arrays obtained, each array is scanned twice, once to form an image of signals corresponding to the marker proteins, and once to form an image of signals corresponding to the sample proteins. The signals and the scanning methods are chosen such that the marker protein spots are separately determinable from the sample protein spots, either directly from the scans themselves or upon manipulating the scans by combining them, subtracting one from the other, or similar procedures. The two images that are thus formed are then processed together in a manner that will use the marker protein signals as landmarks for identification of and correspondence between the sample proteins among the various electropherograms.

As in the manual landmarking of the prior art, the user must add a sufficient number of marker proteins to achieve successful matching of the protein spots among the various electropherograms in a matchset. Nevertheless, the use of marker proteins as a substitute for manual landmarking presents various advantages over the prior art. One advantage is the elimination of the need to select and assign a set of landmark spots each time a new matchset is to be analyzed. Another is the elimination of the need to manually select landmark spots from among the full array of protein spots in an actual sample. A further advantage is the elimination of the requirement that the landmark spots be isolated from other spots so that they can be spatially distinguished on the electropherogram. A still further advantage is the reduction or elimination of errors in landmark selection from one gel to another.

These and other objects, features and advantages of the invention are explained in more detail in the following portions of this specification.

DETAILED DESCRIPTION OF THE INVENTION AND SPECIFIC EMBODIMENTS

Proteins suitable for use as marker proteins can be any proteins that will migrate under the two-dimensional electrophoresis conditions used in the separations to form protein spots in the resulting arrays, and that will not interfere with the migration of the sample proteins. Preferred marker proteins are those that migrate to locations in the array that are well separated from each other and widely dispersed both horizontally and vertically, and yet are close to the locations to which the sample proteins migrate. Particularly preferred marker proteins are those whose spots in the two-dimensional array will appear at locations near the corners of the array, and in or near regions where there is a high concentration of sample protein spots. Marker proteins selected with these considerations will form spots that cover a range of the separation parameters that are used for identification of the sample proteins. Thus, the selection of marker proteins will be governed in many cases by the particular separation parameter(s) used in the two-dimensional electrophoresis. As noted below, the separation parameter(s) and the manner in which the second dimension separation differs from the first dimension separation can vary widely. Two of the most prominent and widely used separation parameters are molecular weight, isoelectric point, and a combination of both, and accordingly, these are preferred separation parameters in the practice of this invention.

The marker proteins may be identical to some of the sample proteins except labeled in a distinguishable manner. Alternatively, the marker proteins may be distinct from the sample proteins. Note that the term "marker proteins" is used herein in the specification and claims to denote the proteins whose spots that will serve as landmarks, and the term "sample proteins" is used herein in the specification and claims to denote the proteins that were present in the sample(s) before the marker proteins were introduced.

The number of marker proteins and the particular proteins to be used as marker proteins are not critical to this invention and may vary widely. Considerations governing the number and choice of marker proteins are the same as those governing the number and choice of reference spots used as landmarks according to procedures of the prior art. Thus, any number can be used that is sufficient to achieve successful matching of sample protein spots among each of the various gels in a matchset. Preferably, at least five marker proteins will be used, more preferably from 5 to 25, and most preferably from 9 to 16. The optimal number of marker proteins for any given gel will be the number that

provides good mapping that will compensate for inter-gel distortion, and yet still allow unambiguous automatic matching of marker spots.

Labeling of the marker proteins can be achieved by any labeling technique that will permit detection in a manner that is selective relative to the sample proteins. Many labeling methods are known and can be used for the marker proteins. Examples of conventional labels are fluorescent labels, chemiluminescent labels, colorimetric labels, and radioisotope labels.

Differentiation between the marker proteins and the sample proteins can be achieved by labeling the sample proteins with labels that are distinguishable from the labels of the marker proteins, or by labeling only one of the two groups of proteins and staining the gels containing the two-dimensional protein spot arrays in a manner that will respond to the marker proteins differently than to the sample proteins. In general, signal differentiation can be achieved by having the two groups of proteins are labeled with different labels, or by labeling one and not the other.

Examples of labels that can be differentiated from each other are fluorescent dyes with different emission maxima. Differential imaging of the two fluorescent dyes can then be performed by focusing on the maxima. Examples of similar fluorescent dyes with different maxima are the cyanine dyes Cy3™ and Cy5.5™, both products of Amersham Life Science, Inc., Arlington Heights, Illinois, USA. Other cyanine dyes can be used, as well as fluorescein dyes, rhodamine dyes such as TAMRA (N,N,N',N'-tetramethyl-6-carboxyrhodamine) and Alexa dyes (sulfonated rhodamines), and phenanthridine dyes such as Texas Red. Stains that can be used with these dyes include SYPRO Ruby, SYPRO Rose, and SYPRO Orange. Examples of other stains are silver stain, copper stain, zinc stain, and Coomassie Brilliant Blue. All of these materials and additional materials for use as further alternatives are available from commercial suppliers, such as Molecular Probes, Inc., Eugene, Oregon, USA. Differentiable radioactive labeling can be achieved, for example, by using different radioisotopes whose signals can be distinguished from each other. Thus, ^{32}P can be used as the label for one group of proteins and ^{35}S as the label for the other. The signals can be separated using an attenuation material such as a thin copper layer. Other combinations will be readily apparent to those skilled in radioisotope labeling. Differentiation can also be obtained by using different types of labels, such as a fluorescent label on one group and staining the

other group with silver stain. Other combinations and methods of differentiation will be readily apparent to those skilled in the art of protein labeling and detection.

The combining of the marker protein spots with the sample protein spots in the two-dimensional electropherograms may be achieved by simply adding the marker proteins to the samples prior to electrophoresis. Labeling of the marker proteins, sample proteins, or both can be performed prior to adding the marker proteins to the samples. In certain applications, it is preferable to leave the sample proteins unlabeled and to thus add the labeled marker proteins to unlabeled sample proteins. After electrophoresis, the gel can be stained with a fluorescent dye that is distinguishable from the fluorescent label that was pre-attached to the marker proteins. By not pre-attaching a label to the sample proteins, one avoids the risk of affecting the electrophoretic migration of the sample proteins due to changes in the mass and/or isoelectric point.

The amounts of marker proteins can vary and are not critical to the invention, but optimal results will generally be obtained if the amounts are selected such that clear detection of both the marker proteins and the sample proteins can be achieved. The marker proteins will typically be added to the samples to concentrations ranging from about 5 nanograms to about 500 nanograms for each marker protein. Differences in signal intensities between the two groups of protein spots can be accommodated by adjustments in the scanning techniques.

The two-dimensional electrophoresis portion of the method of this invention can be performed according to known procedures, which may vary widely. In a typical procedure, the sample is first given a linear separation in an elongate or rod-shaped gel, with an electric potential imposed along the length of the gel. Migration and separation thus occur along the gel axis until the proteins in the sample are distributed among zones positioned along the length of the gel. This is followed by placement of the elongate gel along one edge of a slab gel, and the imposition of an electric potential in a direction lying within the plane of the slab gel and perpendicular to the edge where the elongate gel is placed. The proteins from each zone of the elongate gel migrate into the slab gel in the direction transverse to the axis of the elongate gel. The result is a two-dimensional array of protein spots in the slab gel. By using one mode of separation or one set of separation conditions in the first dimension (the elongate gel) and a different mode or set of separation conditions in the second dimension (the slab gel), highly effective separations can be obtained. For example, the first mode may be one based on charge, such as isoelectric focusing, and the second may be based on molecular weight.

Alternatively, the two dimensions of the separation may be based on the same parameter but may differ in the compositions of the two gels. The two gels may differ in gel concentration or chemical components. As a further alternative, the two dimensions of the separation may be based on the same parameter and performed in gels of the same composition and concentration, but differ in a separation condition, such as a stepwise difference in pH, for example. Still further alternatives are the use of a homogeneous gel in one of the two dimensions and a gradient gel in the other, the use of two different protein solubilizers in the two dimensions, or two different concentrations of the same protein solubilizer, and the use of a nonchanging buffer system in one dimension and a changing (gradient, for example) buffer system in another dimension.

Once the separations have been performed and the two-dimensional arrays of protein spots, representing both marker proteins and sample proteins, have been established in the gels, each gel is scanned twice, once to obtain an image of the marker protein spots and once to obtain an image of the sample protein spots. The order of the scans is not critical; either one can be performed first. The two scans are performed in such a manner that they permit differentiation of the marker protein spots from the sample protein spots. This can be achieved by selectively scanning the marker protein spots in one scan and selectively scanning the sample protein spots in the other scan, or selectively scanning either the marker protein spots or the sample protein spots in one scan, scanning all spots in the other scan, and subtracting the signals from the first scan from those of the second scan. Preferably, each of the two scans is a selective scan, one detecting only the marker protein spots and the other only the sample protein spots.

This invention is directed to two-dimensional electrophoresis in general, although the most common separation media, and those that are preferred in the practice of this invention, are gels. The type and composition of the gels used for either of the two dimensions are not critical to the invention and may vary. Agarose gels and polyacrylamide gels are two prominent and common examples, and analogs of these gels such as formamide and urea gels can also be used. The preferred gels are polyacrylamide gels.

Scanning of the gels can be achieved by known techniques with commercially available scanning equipment. Examples of such equipment are multi-color imaging charge coupled devices and laser imaging systems. Other systems will be readily apparent to those skilled in the art. Selection of the appropriate scanning system

will depend on the types of labels used and the manner in which the marker protein signals are differentiated from the sample protein signals.

The scanning data can be processed into digital form by conventional computer software of the type that is currently used for processing data of this kind, such as the various commercially available programs listed above or simple variations on these programs. The two scans for each electropherogram in the matchset can be processed together by methods that use the marker protein data as landmarks and achieve the same or equivalent results as the prior art systems which rely on landmarking. Using conventional algorithms, the data can be processed to identify individual proteins according to the parameters used as bases for the separations, match corresponding protein data among the various gels in the matchset, and compile and present the progression or other changes of the amounts of each protein throughout the entire sequence of gels. If desired, further identification of the proteins can be achieved by MALDI-TOF (Matrix-Assisted Laser Desorption Ionization — Time of Flight) analysis.

Thus, by the practice of this invention, a series of samples, containing from as little as a few samples up to several hundred samples, can be analyzed and compared, and minute, progressive changes in many distinct proteins present in each of the samples can be observed and recorded simultaneously.

The foregoing is offered primarily for purposes of illustration. It will be readily apparent to those skilled in the art of two-dimensional electrophoresis and protein analysis that further modifications, variations, and substitutions can be made without departing from the spirit and scope of the invention.

WHAT IS CLAIMED IS:

1. A method for the detection and quantification of proteins in a plurality of samples and for the comparison of amounts of such proteins, defined as sample proteins, between individual samples in said plurality of samples, said method comprising:

(a) combining each of said samples with a plurality of marker proteins each having a known identity and known characteristics that affect its migration in two-dimensional electrophoresis, said marker proteins each being labeled in a manner that is distinguishable from said sample proteins;

(b) performing two-dimensional electrophoresis on said samples thus combined with said marker proteins to form a two-dimensional array of protein spots for each sample, each said array including marker protein spots corresponding to said marker proteins and sample protein spots corresponding to said sample proteins;

(c) scanning each array to form a first image of signals corresponding to said marker protein spots and a second image of signals corresponding to said sample protein spots, the signals of said first image distinguishable from those of said first image; and

(d) processing said first and second images together for each of said arrays, using said marker protein spots to automatically match corresponding sample protein spots among individual arrays, and comparing corresponding sample protein spots, thus matched, among individual arrays to determine differences between individual samples in the amounts of said sample proteins.

2. A method in accordance with claim 1 in which said two-dimensional electrophoresis is two-dimensional polyacrylamide gel electrophoresis.

3. A method in accordance with claim 1 in which said known characteristics of said marker proteins are their molecular weights, and step (d) further comprises assigning a molecular weight to each of said sample protein spots based on the molecular weights of said marker proteins.

4. A method in accordance with claim 1 in which said known characteristics of said marker proteins are their isoelectric points, and step (d) further

comprises assigning an isoelectric point to each of said sample protein spots based on the isoelectric points of said marker proteins.

5. A method in accordance with claim 1 in which said known characteristics of said marker proteins are their molecular weights and isoelectric points, and step (d) further comprises assigning a molecular weight and an isoelectric point to each of said sample protein spots based on the molecular weights and isoelectric points of said marker proteins.

6. A method in accordance with claim 1 in which said two-dimensional electrophoresis is two-dimensional polyacrylamide gel electrophoresis and said known characteristics of said marker proteins are molecular weight and isoelectric point.

7. A method in accordance with claim 1 in which said plurality of marker proteins consists of from 9 to 16 marker proteins.

8. A method in accordance with claim 1 further comprising labeling said sample proteins with a first radioisotope and said marker proteins with a second radioisotope that is different from said first radioisotope and distinguishable by rasioisotope detection means.

9. A method in accordance with claim 1 further comprising labeling each of said sample proteins with a common type of label and each of said marker proteins with a common type of label distinguishable from that of said sample proteins, one of said types being a fluorescent label and the other being a colorimetric stain.

10. A method in accordance with claim 1 further comprising labeling said sample proteins with a first fluorescent dye and said marker proteins with a second fluorescent dye distinguishable from said first fluorescent dye.